Report to SNF, Inc.

Acrylamide

(CAS 000079-06-1)

Development of BMD for Reproductive Endpoints

By

Bruce Allen

Environ International

Analysis of Reproductive Toxicity Data

We conducted a BMD analysis of the reproductive toxicity data from Chapin et al. (1995) and Tyl et al. (2000). The endpoints selected for evaluation were the ones for which those authors indicated that one or more of the dosed groups had responses significantly different from the control responses. However, in some cases, we did not analyze both of two endpoints that were essentially reparameterizations of the same effect (i.e., we modeled average number of live fetuses per dam but not total post-implantation loss, since those two endpoints are essentially measuring the same effect when there is no difference in average number of implantations).

The endpoints of interest with respect reproductive toxicity were quantitatively summarized in the published reports by group means and measures of variability (standard errors or standard deviations). Even though the data would actually be dichotomous and nested (e.g., counts of resorptions within litter, within dose group) the treatment of the litter averages as continuous variables is adequate and amenable to BMD estimation (Allen et al., 1995). The models used to represent the dose-response behavior of those continuous endpoints are those implemented in EPA's Benchmark Dose Software (BMDS Version 1.3.1, U.S. EPA, 2001).

For this analysis, the BMDS models implemented were the power model and the linear model. The polynomial model was not run because it would not run on computers running newer versions of Windows. The Hill model was not run because experience has shown that that model also has difficulty running on data sets having 4 or fewer dose groups (there were only four groups in each of the two studies analyzed).

The power model is represented by the equation

$$\mu(d) = \gamma + \beta d^{\alpha}$$

where $\mu(d)$ indicates the mean of the response variable following exposure to dose d. The parameter α was restricted to be greater than or equal to 1. The linear model was obtained when α was fixed at a value of 1.

In the case of continuous endpoints, one must assume something about the distribution of individual observations around the dose-specific mean values defined by the above models. The assumptions imposed by BMDS were used in this analysis: individual observations were assumed to vary normally around the means with variances given by the following equation:

$$\sigma_i^2 = \sigma^2 \cdot [\mu(d_i)]^{\rho}$$

where both σ^2 and ρ were parameters estimated by the model.

Given those assumptions about variation around the means, maximum likelihood methods were applied to estimate all of the parameters, where the log-likelihood to be maximized is (except for an additive constant) given by

$$L = \Sigma\, [(N_i/2)\cdot\ln(\sigma_i^2) + (N_i - 1)s_i^2/2\sigma_i^2 + N_i\{m_i - \mu(d_i)\}^2/2\sigma_i^2]$$

where N_i is the number of individuals in group i exposed to dose d_i, and m_i and s_i are the observed mean and standard deviation for that group. The summation runs over i from 1 to k (the number of dose groups).

The goodness-of-fit statistics produced by BMDS for the power model are based on likelihood ratio statistics. These statistics look at the differences in log-likelihoods produced by different models, and it can be determined if one model does a "significantly better" job in fitting the data than another model. In particular, the test statistic examined here compares the maximized log-likelihood associated with the fitted model to the log-likelihood maximized with each dose group considered to have a mean completely independent of the means of the other dose groups, and the variances modeled as shown above. It is always the case that the latter log-likelihood will be at least as great as the model-associated log-likelihood, but if the model of the change in mean does a "reasonable" job of fitting the data, the difference between the two log-likelihoods will not be too great. A formal statistical test reflecting this idea uses the fact that twice the difference in the log-likelihoods is distributed as a chi-square random variable. The

degrees of freedom associated with that chi-squared test statistic are equal to the difference between the number of parameters fit by the model (including the parameters σ^2 and ρ defining how variances change as a function of mean response level) and the number of dose groups plus 2 (which is equal to the number of parameters estimated by the "model" assuming independence of dose group means but with variances defined as above. We also examined the impact of assuming a constant variance and chose the constant variance model if its fit was satisfactory.

For the standard approach to BMR definition for the continuous models, BMDs were implicitly defined as follows:

$$\left| \mu(BMD) - \mu(0) \right| = \delta \cdot \sigma_1$$

where σ_1 is the model-estimated standard deviation in the control group. In other words, the BMR was defined as a change in mean corresponding to some multiplicative factor of the control group standard deviation. The value of δ used in this analysis was 1.1. This value was chosen based on the work of Crump (1995), who showed that that choice corresponded to an additional risk of 10% when the background response rate was assumed to be 1%, with normal variation around the means (and constant standard deviation). Although the current analyses allowed for nonconstant standard deviations, the value of 1.1 was used for two reasons. First, the difference between additional and extra risk is small when the background rate is 1% or less, so that the change from additional to extra risk will have minimal impact on the correspondences proven by

Crump (1995). Second, there can be no such generic, a priori correspondences when standard deviations are allowed to vary in a manner determined only after the model fitting is accomplished. Thus, to avoid data set- and model-specific choices for δ, the correspondences proven by Crump (1995) can be used as the best available, consistent definition of the benchmark response. The definition of the BMR as a change in mean of 1.1 times the control standard deviation is very close to the default value of 1 standard deviation recommended by recent draft EPA guidelines (U.S. EPA, 2000).

The reproductive endpoints selected for the BMD analysis were the following:

Chapin et al. (1995) – Live pups/litter from matings of the F0 mice; early resorptions/litter and live fetuses/litter from their dominant lethal segment; and live pups/litter from matings of the F1 mice.

Tyl et al. (2000) – Implantations/litter and live pups/litter from mating of F0 rats; implantations/litter and live implants/litter from their dominant lethal segment; and implantations/litter and live pups/litter from matings of the F1 rats.

Results of those analyses are shown in the following tables:

Table 1: Summary of BMD Analysis of Chapin et al. (1995) Reproductive Endpoints

Endpoint	Model	Goodness-of-fit p-value	BMD	BMDL
F0, Live pups/litter	Power	0.49*	14	7.3
	Linear	0.49*	14	7.3
Dominant lethal, early resorptions/litter	Power	0.07*	7.2	5.2
	Linear	0.10*	6.9	4.7
Dominant	Power	0.14	7.3	6.6

lethal, Live fetuses/litter	Linear	0.02	9.0	5.6
F1, Live pups/litter	Power	0.01	5.9	2.8
	Linear	<.00001	NA	NA

Notes: The results shown are for the versions of the models with the most appropriate handling of the variances. If BMDS suggested that a nonconstant variance model was needed, then the results are those for the nonconstant variance version; otherwise they are for the constant variance version. Goodness-of-fit p-values are relative to the independent-means model having the same treatment of the variances. An asterisk on the p-value indicates that the modeled variances were not in good agreement with the observed variances. The BMD and BMDL are the maximum likelihood and 95% lower bound estimates of the dose corresponding to a change in mean response equal to 1.1 times the control group standard deviation. "NA" indicates that the model was not able to estimate the BMD and BMDL values because it resulted in a flat curve.

Table 2: Summary of BMD Analysis of Tyl et al. (2000) Reproductive Endpoints

Endpoint	Model	Goodness-of-fit p-value	BMD	BMDL
F0, Implantations/litter	Power	0.40*	5.0	4.1
	Linear	0.08*	4.5	3.3
F0, Live pups/litter	Power	1.0	4.4	3.5
	Linear	0.01	3.1	2.5
Dominant lethal, Implantations/litter	Power	0.62	5.2	5.0
	Linear	0.34	9.6	5.7
Dominant lethal, Live implants/litter	Power	0.74	5.1	4.8
	Linear	0.20	6.6	4.5
F1, Implantations/litter	Power	0.0002*	4.9	3.5
	Linear	0.0002*	3.4	2.6
F1, live pups/litter	Power	0.07*	3.3	2.1
	Linear	<.00001	NA	NA

Notes: The results shown are for the versions of the models with the most appropriate handling of the variances. If BMDS suggested that a nonconstant variance model was needed, then the results are those for the nonconstant variance version; otherwise they are for the constant variance version. Goodness-of-fit p-values are relative to the independent-means model having the same treatment of the variances. An asterisk on the p-value indicates that the modeled variances were not in good agreement with the observed variances. The BMD and BMDL are the maximum likelihood and 95% lower bound estimates of the dose corresponding to a change in mean response equal to 1.1 times the control group standard deviation. "NA" indicates that the model was not able to estimate the BMD and BMDL values because it resulted in a flat curve.

The results shown above are relatively consistent across the two studies and across endpoints. The linear models were consistently not as good as the power models in replicating the dose-response observations. This was because the data sets analyzed had a particular pattern in common: very little change from control values for the lowest two dosed groups followed by a significant change for the highest dose group. Such patterns tend to require more curvature in the dose-response than is available with a linear model. If a linear model is forced, then the estimated control-group mean tends to overestimate the observed value so that the constant slope of the curve does not entail too much underestimation of the responses in the two lowest dosed groups. Despite the occasional difficulty the models had in matching the observed variances, the power model did tend to fit the mean responses well. Thus, notwithstanding the fit issues associated with modeling of the variances, the range of BMDL values from the above tables (2.1 to 7.3 mg/kg/day) can be taken to be indicative of the sensitivity of these reproductive parameters to ACR exposure. The results of the Tyl et al. (2000) study may be slightly better estimates to use as starting points, because that study had somewhat larger sample sizes and slightly lower doses than did the Chapin et al. (1995) study. The differences in BMD and BMDL estimates between comparable endpoints for the F0 and F1 generations were not great.